

# Continuity properties of a factor of Markov chains

Walter A. F. de Carvalho<sup>1</sup>, Sandro Gallo<sup>2</sup>, and Nancy L. Garcia<sup>1\*</sup>

September 23, 2014

## Abstract

Starting from a Markov chain with a finite alphabet, we consider the chain obtained when all but one symbol are undistinguishable for the practitioner. We study necessary and sufficient conditions for this chain to have continuous transition probabilities with respect to the past.

## 1 Introduction

Consider a positive recurrent, stationary Markov chain  $\mathbf{X} = \{X_t, t \in \mathbb{Z}\}$  in a finite (or countable) alphabet  $A$  with transition matrix  $P^{\mathbf{X}}$  and a factor map,  $\pi : A \rightarrow B$  with  $\text{card}(B) < \text{card}(A)$ . Hidden Markov models obtained through factor mappings of Markov Chains have received a lot of attention for a long time and an exhaustive listing would be impossible. We therefore mention only some paper that are most directly related to our interest. Harris (1955) proved that, in general, the image process  $\mathbf{Y} = \{Y_t = \pi(X_t), t \in \mathbb{Z}\}$  is a chain with infinite order with continuous transition probabilities decaying exponentially fast if  $P$  is strictly positive (i.e.,  $P_{ij}^{\mathbf{X}} > 0$  for all  $i, j \in A$ ). This result was then extended in Verbitskiy (2011a) and Verbitskiy (2011b), to the case where the original chain  $\mathbf{X}$  is not necessarily a Markov chain, but only needs to be a chain with strictly positive and continuous transition probabilities. In the statistical physics and dynamical systems contexts, Chazottes and Ugalde

---

\*Corresponding author; <sup>1</sup> University of Campinas; <sup>2</sup> Federal University of Rio de Janeiro; Brasil

(2003) were interested in determining whether the image process of a Markov process satisfies (Bowen's) Gibbsianess. Finally, another related study is that of preservation of (order 1) Markovianess, we refer for instance to Burke and Rosenblatt (1958) who gave necessary and sufficient conditions under which  $\mathbf{Y}$  is still a Markov process.

In this work, we focus our efforts on the special case where  $B = \{0, 1\}$  and  $\pi^{-1}(1)$  is a singleton. Without loss of generality, we take  $A = \{1, \dots, m\}$  and  $\pi(x) = \mathbf{1}(x = 1)$ . In this case, it is well-known that the image process  $\mathbf{Y} = \{Y_t, t \in \mathbb{Z}\}$  is always a discrete time binary renewal chain, and it is an easy matter to obtain an explicit matricial form for its transition probabilities. This explicitness then allows us to search not only for sufficient conditions (as in most of the previously cited article), but also for necessary conditions under which the image process has continuous transition probabilities. We refer the reader to Section 3 for a more detailed explanation of our objectives, assumptions and results.

The paper is organized as follows. In Section 2 we introduce the notation and basic definitions. The main results and their proofs are given in Section 3, each theorem being directly followed by its proof. We conclude with a section containing some examples and a discussion comparing our results to those of the related literature.

## 2 Notation and basic definitions

### Notation.

Let  $A$  be a countable alphabet. Given two integers  $m \leq n$ , we denote by  $a_m^n$  the string  $a_m \dots a_n$  of symbols in  $A$ . For any  $m \leq n$ , the length of the string  $a_m^n$  is denoted by  $|a_m^n|$  and defined by  $n - m + 1$ . Given two strings  $v$  and  $v'$ , we denote by  $vv'$  the string of length  $|v| + |v'|$  obtained by concatenating the two strings. The concatenation of strings is also extended to the case where  $v = \dots a_{-2}a_{-1}$  is a semi-infinite sequence of symbols. Let  $A^{-\mathbb{N}} = A^{\{\dots, -2, -1\}}$  be the set of all infinite strings of past symbols. Finally, we denote by  $\underline{a} = \dots a_{-2}a_{-1}$  the elements of  $A^{-\mathbb{N}}$ .

### Kernel, compatibility and continuity.

**Definition 2.1** (Kernel). *A transition probability kernel (or simply kernel in*

the sequel) on a countable alphabet  $A$  is a (measurable) function

$$\begin{aligned} P : A \times A^{-\mathbb{N}} &\rightarrow [0, 1] \\ (a, \underline{a}) &\mapsto P(a|\underline{a}) \end{aligned} \tag{1}$$

such that  $\sum_{a \in A} P(a|\underline{a}) = 1$  for any  $\underline{a} \in A^{-\mathbb{N}}$ .

For instance, a given kernel  $P$  is Markovian of order  $k \geq 1$  if  $P(a|\underline{a}) = P(a|\underline{b})$  for any  $\underline{a}$  and  $\underline{b}$  such that  $a_{-k}^{-1} = b_{-k}^{-1}$ . So the above definition is a natural extension of the notion of transition matrix, the later being a terminology commonly used for  $k$ -steps Markov chains.

A kernel is sometimes called set of transition probabilities, a terminology that we used in the introduction. It is important to notice that these are transition probabilities with respect to the “past”. When considering transition probabilities with respect to both, past and future, as we will do in Section 4 when comparing our results to the notion of Gibbsianess, we will use the terminology *specification*.

In this paper, stochastic chains are regarded as bi-infinite sequence of random variables whose dependence information is given by the kernel. We therefore need to define what we mean by a stochastic chain being compatible with a given kernel.

**Definition 2.2** (Compatibility). *A stationary stochastic chain  $\mathbf{X} = (X_n)_{n \in \mathbb{Z}}$  is said to be compatible with a kernel  $P$  if the later is a regular version of the conditional probabilities of the former, that is*

$$\mathbb{P}(X_0 = a_0 | X_{-\infty}^{-1} = a_{-\infty}^{-1}) = P(a|a_{-\infty}^{-1}) \tag{2}$$

for every  $a_0 \in A$  and  $\mathbb{P}$ -a.e.  $a_{-\infty}^{-1}$  in  $A^{-\mathbb{N}}$ .

Let us emphasize that stationarity here means that for any  $n \geq 1$  and any  $a_0^{n-1}$ ,  $\mathbb{P}(X_0^{n-1} = a_0^{n-1}) = \mathbb{P}(X_k^{k+n-1} = a_0^{n-1})$  for any  $k \in \mathbb{Z}$ .

As an example for Definition 2.2, a stationary 1-step Markov chain satisfies that  $\mathbb{P}(X_0 = a_0 | X_{-1} = a_{-1}) = P_{a_{-1}a_0}$  for any  $a_{-1}, a_0 \in A$ , where  $P_{a_{-1}a_0} := P(a_0|\underline{a})$ , since this later does not depend on  $a_{-\infty}^{-2}$ . In what follows,  $P_{ij}$ ,  $i, j \in A$ , will always denote the entries of a transition matrix on  $A$ .

We finally need to introduce what we mean by continuous kernel, the main notion of interest of the present paper.

**Definition 2.3** (Continuity). *A kernel  $P$  is continuous (with respect to the product topology) at some point  $\underline{a}$  if  $P(a|a_{-i}^{-1}\underline{z}) \rightarrow P(a|\underline{a})$  whenever  $i$  diverges, for any  $\underline{z}$ . Naturally, a kernel  $P$  is continuous if it is continuous at every point.*

In the above definition,  $P(a|a_{-i}^{-1}\underline{z})$  stands for the transition probability from the (concatenated) past  $\underline{z}a_{-i}^{-1}$  to the symbol  $a$ , we reversed time in the conditioning.

It is clear that  $k$ -steps Markov kernels are continuous since for them,  $P(a|a_{-i}^{-1}\underline{z}) = P(a|a_{-k}^{-1}\underline{z})$  for any  $i \geq k$ . So continuous kernels constitute a natural extension of  $k$ -steps Markov kernels. The next definition aims to quantify *how much* continuous is a given kernel.

**Definition 2.4** (Continuity rate). *The continuity rate of a kernel  $P$  is defined, for any  $k \geq 1$ , by*

$$\beta(k) := \sup_{a_{-k}^0} \sup_{\underline{b}, \underline{c}} |P(a_0|a_{-k}^{-1}\underline{b}) - P(a_0|a_{-k}^{-1}\underline{c})|.$$

Observe that if the alphabet  $A$  is finite, the compactness of  $A^{-\mathbb{N}}$  implies that  $P$  is continuous if and only if its continuity rate converges to 0.

### 3 Results and Proofs

One of the main assumption of the present paper is that we restrict our considerations to a particular case of mapping.

**Definition 3.1** (Aggregation map). *A map  $\pi : A \rightarrow \{0, 1\}$  defined by  $\pi(a) = \mathbf{1}(a = 1)$  will be called aggregation map.*

**Notation alert 3.1.** *We will use the same notation  $\pi$  to denote the map  $\pi : A^n \rightarrow \{0, 1\}^n$  defined by  $\pi(a_1^n)_i = \mathbf{1}(a_i = 1)$ . The generalization for infinite sequences and processes is analogous.*

In the sequel, we will essentially study two stochastic chains, the original Markov chain  $\mathbf{X}$  and its image  $\mathbf{Y}$ . in order to avoid confusions, we will index quantities by the related stochastic chains:  $P^{\mathbf{X}}$  and  $P^{\mathbf{Y}}$  for the kernels, and  $\beta^{\mathbf{X}}(k)$  and  $\beta^{\mathbf{Y}}$  for the continuity rates.

### 3.1 Preliminary considerations

Focussing on aggregation maps (see Definition 3.1) allows to obtain a closed formula for the continuity rate of the image process, as stated by the next proposition.

For any  $\underline{a} \in \{0, 1\}^{-\mathbb{N}}$ , let

$$\ell(\underline{a}) := \inf\{i \geq 0 : a_{-i-1} = 1\},$$

using the convention that  $\ell(\dots 00) = \infty$ . Also let us introduce the  $[0, 1]$ -valued sequence  $p_k$ ,  $k \geq 0$ , defined by  $p_0 := P_{11} = \mathbb{P}(X_0 = 1 | X_{-1} = 1)$  and for any  $k \geq 1$ ,

$$p_k := \frac{\sum_{b_{-k}^{-1} \in \pi^{-1}(0_{-k}^{-1})} \mathbb{P}(X_{-k}^{-1} = b_{-k}^{-1}, X_0 = 1 | X_{-k-1} = 1)}{\sum_{b_{-k}^{-1} \in \pi^{-1}(0_{-k}^{-1})} \mathbb{P}(X_{-k}^{-1} = b_{-k}^{-1} | X_{-k-1} = 1)}. \quad (3)$$

**Proposition 3.1.** *If  $\mathbf{X}$  is a stationary positive recurrent Markov chain, then the continuity rate of  $P^{\mathbf{Y}}$  is*

$$\beta^{\mathbf{Y}}(k) \leq \sup_{l, m \geq k} |p_l - p_m|.$$

*Proof.* For all  $k \geq 0$ , we will compute, for any  $i \geq 1$  and any  $a_{-i}^{-1} \in \{0, 1\}^i$ ,

$$\mathbb{P}(Y_0 = 1 | Y_{-i-k-1}^{-k-2} = a_{-i}^{-1}, Y_{-k-1} = 1, Y_{-k}^{-1} = 0_{-k}^{-1}). \quad (4)$$

It equals, by the definition of conditioning

$$\frac{\mathbb{P}(Y_{-i-k-1}^{-k-2} = a_{-i}^{-1}, Y_{-k-1} = 1, Y_{-k}^{-1} = 0_{-k}^{-1}, Y_0 = 1)}{\mathbb{P}(Y_{-i-k-1}^{-k-2} = a_{-i}^{-1}, Y_{-k-1} = 1, Y_{-k}^{-1} = 0_{-k}^{-1})} \quad (5)$$

Now, using that  $\pi$  is the aggregating function of Definition 3.1 and that  $\mathbf{X}$  is Markovian, we have that the numerator of (5) equals

$$\sum_{c_{-i}^{-1} \in \pi^{-1}(a_{-i}^{-1})} \mathbb{P}(X_{-i-k-1}^{-k-2} = c_{-i}^{-1}, X_{-k-1} = 1) \sum_{b_{-k}^{-1} \in \pi^{-1}(0_{-k}^{-1})} \mathbb{P}(X_{-k}^{-1} = b_{-k}^{-1}, X_0 = 1 | X_{-k-1} = 1),$$

and analogously, the denominator equals

$$\sum_{c_{-i}^{-1} \in \pi^{-1}(a_{-i}^{-1})} \mathbb{P}(X_{-i-k-1}^{-k-2} = c_{-i}^{-1}, X_{-k-1} = 1) \sum_{b_{-k}^{-1} \in \pi^{-1}(0_{-k}^{-1})} \mathbb{P}(X_{-k}^{-1} = b_{-k}^{-1} | X_{-k-1} = 1).$$

Thus, the fraction in (5) factorizes and we obtain that (4) equals  $p_k$ , independently of  $a_{-k}^{-1}$ . This in particular means that, for any  $\underline{a} \in \{0, 1\}^{-\mathbb{N}} \setminus \{\dots 00\}$ ,  $\mathbb{P}(Y_0 = 1 | Y_{-k}^{-1} = a_{-k}^{-1})$  converges, since it is constant for any  $k \geq \ell(\underline{a}) + 1$ . By Kalikow (1990),  $\mathbf{Y}$  has a continuous kernel  $P^{\mathbf{Y}}$  if and only if  $\mathbb{P}(Y_0 = 1 | Y_{-k}^{-1} = a_{-k}^{-1})$  converges uniformly on  $\{0, 1\}^{-\mathbb{N}}$ , which in the present case, due to compactness of  $\{0, 1\}^{-\mathbb{N}}$ , amounts to say that  $P^{\mathbf{Y}}$  is continuous if and only if  $\mathbb{P}(Y_0 = 1 | Y_{-k}^{-1} = a_{-k}^{-1})$  converges for any  $\underline{a} \in \{0, 1\}^{-\mathbb{N}}$ . Since this quantity converges for any  $\underline{a} \in \{0, 1\}^{-\mathbb{N}} \setminus \{\dots 00\}$ , we have continuity if and only if the convergence occurs at  $\dots 00$ .

Now, by the definition of compatibility, we can take  $P^{\mathbf{Y}}$  defined for any  $\underline{a}$  by  $P^{\mathbf{Y}}(1 | \underline{a}) = \lim_k \mathbb{P}(Y_0 = 1 | Y_{-k}^{-1} = a_{-k}^{-1})$ . In other words, these limits can be taken to be the regular version of the conditional probabilities. In particular,  $P^{\mathbf{Y}}(1 | \underline{a}) = p_{\ell(\underline{a})}$  for any  $\underline{a} \in \{0, 1\}^{-\mathbb{N}} \setminus \{\dots 00\}$  and we can take  $P^{\mathbf{Y}}(1 | \dots 00) = \lim_k p_k =: p_\infty$  to preserve continuity.

Then, we can compute the continuity rate of  $P^{\mathbf{Y}}$ :

$$\begin{aligned} \beta^{\mathbf{Y}}(k) &:= \sup_{a_{-k}^{-1}} \sup_{\underline{b}, \underline{c}} |P^{\mathbf{Y}}(1 | a_{-k}^{-1} \underline{b}) - P^{\mathbf{Y}}(1 | a_{-k}^{-1} \underline{c})| \\ &= \sup_{\underline{b}, \underline{c}} |P^{\mathbf{Y}}(1 | 0_{-k}^{-1} \underline{b}) - P^{\mathbf{Y}}(1 | 0_{-k}^{-1} \underline{c})| \\ &= \sup_{\underline{b}, \underline{c}} |p_{\ell(\underline{b})+k} - p_{\ell(\underline{c})+k}| \\ &= \sup_{l, m \geq k} |p_l - p_m| \end{aligned}$$

where in the second line we used the fact that if  $a_{-i} = 1$  for some  $i \in \{1, \dots, k\}$ , then as we said above  $P^{\mathbf{Y}}(1 | a_{-k}^{-1} \underline{b}) = P^{\mathbf{Y}}(1 | a_{-k}^{-1} \underline{c}) = p_i$ , and in the third line, we used the fact that the transition probability  $P^{\mathbf{Y}}(1 | \underline{a})$  only depends on  $\ell(\underline{a})$ .  $\square$

In the above argument, observe that the value of  $P^{\mathbf{Y}}(1 | \dots 00)$  can be chosen arbitrarily. For instance, we could have taken a different value than the limit of  $p_k$ , creating a discontinuity at  $\dots 00$  according to Definition 2.3. But this is only an artificial discontinuity, because the regular version  $P^{\mathbf{Y}}$  is defined up to sets of null measure. Thus we will always choose the continuous one.

In certain cases however, it is impossible to make  $P^{\mathbf{Y}}$  continuous just changing its value on a set of null measure. For instance, let  $\mathbf{X}$  be a Markov chain with alphabet  $A = \{1, 2, 3\}$  and transition matrix given by

$$P^{\mathbf{X}} = \begin{pmatrix} \alpha & 1 - \alpha & 0 \\ \beta & 0 & 1 - \beta \\ \gamma & 1 - \gamma & 0 \end{pmatrix}.$$

Let  $\pi : A \rightarrow \{0, 1\}$  be such that  $\pi(1) = 1$ ,  $\pi(2) = \pi(3) = 0$ . In this case, simple calculations show that  $p_0 = \alpha$ , for  $k \geq 1$  odd,  $p_k = \beta$  and for  $k \geq 2$  even,  $p_k = \gamma$ . What happens then is, borrowing the terminology used in Gibbs/Non-Gibbs literature that the kernel  $P^{\mathbf{Y}}$  has an *essential discontinuity* at the point  $\dots 00$ , independently of the value  $P(1|\dots 00)$  we choose. We refer to Fernández, Gallo and Maillard (2011) for a complete discussion.

The main objective of this paper is to understand which condition for  $P^{\mathbf{X}}$  leads to a continuous/discontinuous process  $\mathbf{Y}$ . To do so, Proposition 3.1 says that it suffices to obtain sufficient conditions for  $p_k$  to converge (or not). To study  $p_k$ , we are going to use a decomposition of the transition matrix  $P^{\mathbf{X}}$  similar to the one used by Darroch and Seneta (1967) in their study of quasi-stationary measures. This decomposition will allow us to use Perron-Forbenius theory.

Recall that  $P^{\mathbf{X}}$  is an  $m \times m$  stochastic matrix. Let

$$P^{\mathbf{X}} = \begin{pmatrix} P_{11} & V \\ W^t & P \end{pmatrix} \quad (6)$$

where

- $V$  is the row vector  $(P_{12} \dots P_{1m})$ ;
- $W$  is the row vector  $(P_{21} \dots P_{m1})$ , and  $W^t$  denotes the transpose of  $W$ ;
- $P$  is the  $(m-1) \times (m-1)$  sub-matrix  $P_{ij}, i, j \in \{2, \dots, m\}$ .

Denote by  $\mathbf{1}$  and  $\mathbf{0}$  the row vectors formed by ones and zeros, respectively. To avoid trivial cases, we will always consider  $V \neq \mathbf{0}$  and  $W \neq \mathbf{0}$ .

**Notation alert 3.2.** *In the sequel,  $P$  will always denote the sub-matrix of  $P^{\mathbf{X}}$ . The kernels will always be indexed by the corresponding process.*

Recall that  $p_0 = P_{11}$ . We can now rewrite  $p_k$ ,  $k \geq 1$  (see (3)) in the following form

$$p_{k+1} = \frac{V P^k W^t}{V P^k \mathbf{1}^t}, \quad \text{if } k \geq 0 \quad (7)$$

with  $P^0 = I$  being the identity  $(m - 1) \times (m - 1)$  matrix.

Before we come to the statements of the results, let us recall the following very basic definitions concerning matrices.

**Definition 3.2.** A square matrix  $(P_{ij})$  is called irreducible if, for any  $i, j$  there exists an  $n(i, j)$  such that  $P_{ij}^n > 0$ . Otherwise, it is called reducible.

**Definition 3.3.** A square matrix  $(P_{ij})$  is called primitive if it is non-negative and there is an  $n > 0$  such that  $P_{ij}^n \neq 0$  for all  $i, j$ .

**Definition 3.4.** Let  $(P_{ij})$  be non-negative. Fix an state  $i$  and define the period of state  $i$  as

$$h(i) := \text{gcm}\{n : P_{ii}^n > 0\}.$$

In the case of irreducible matrices, every state has the same period  $h$  and this number is called the period of  $P$ . If  $h = 1$ ,  $P$  is called aperiodic.

Thus, primitive matrices are irreducible aperiodic non-negative matrices.

### 3.2 The irreducible case

Let  $\mathbf{X}$  be a stationary Markov process with space state  $A = \{1, 2, \dots, m\}$  and its transition matrix  $P^{\mathbf{X}}$  decomposed as (6) with  $P$  being irreducible matrix. Let  $\mathbf{Y} = \pi(\mathbf{X})$  where  $\pi$  is an aggregation map. Theorems 3.1 will give necessary and sufficient conditions under which the process  $\mathbf{Y}$  is continuous as well as its continuity rate.

**Theorem 3.1.** 1. If  $P$  is primitive, then  $P^{\mathbf{Y}}$  is continuous. Moreover

$$\beta^{\mathbf{Y}}(n) = O\left(n^{d_2-1} \left(\frac{|\lambda_2|}{\lambda_1}\right)^n\right)$$

where  $\lambda_1$  is the maximal eigenvalue of  $P$ ,  $\lambda_2$  and  $d_2$  are, respectively, the value and multiplicity of the second largest eigenvalue in absolute value.

2. If  $P$  has period  $h$ , then the process  $\mathbf{Y}$  is continuous if, and only if,

$$\frac{V.G^*.P^r.W^t}{V.G^*.P^r.\mathbf{1}^t} = \frac{V.G^*.W^t}{V.G^*.\mathbf{1}^t} \quad (8)$$

for all  $r = 1, \dots, h - 1$ , where  $G^* = \phi^* \cdot (\psi^*)^t > 0$ , and  $\phi^*$  and  $\psi^*$  are the Perron vectors corresponding to the Perron value  $\lambda_1^*$  from  $P^h$  and  $(P^h)^t$ , respectively.



*Proof.* If  $P$  is **primitive**, then there exists a unique maximal eigenvalue  $\lambda_1$  (Perron-Frobenius eigenvalue or Perron root), which is strictly greater in absolute value than all other eigenvalues and  $\lambda_1$  is the spectral radius of  $P$ . Therefore,

$$\lim_{n \rightarrow \infty} \left( \frac{P}{\lambda_1} \right)^n = G$$

and

$$p_\infty = \lim_{n \rightarrow \infty} p_n = \frac{VGW^t}{VG\mathbf{1}^t} \quad (9)$$

where  $G = \phi.\psi^t > 0$ , and  $\phi$  and  $\psi$  are the Perron vectors corresponding to  $\lambda_1$  from  $P$  and  $P^t$ , respectively.

Moreover, let  $\lambda_2$  be the second largest eigenvalue in absolute value and  $d_2$  its multiplicity. Then, it is known (Theorem 1.2, p. 9, Seneta (2006)) that

$$\left( \frac{P}{\lambda_1} \right)^n = G + O \left( n^{d_2-1} \cdot \left( \frac{|\lambda_2|}{\lambda_1} \right)^n \right). \quad (10)$$

$$\beta^{\mathbf{Y}}(n) = |p_n - p_\infty| = \left| \frac{V.P^n.W^t}{V.P^n.\mathbf{1}^t} - \frac{V.G.W^t}{V.G.\mathbf{1}^t} \right| = O \left( n^{d_2-1} \cdot \left( \frac{|\lambda_2|}{\lambda_1} \right)^n \right).$$

If  $P$  is **not primitive**, without loss of generality we can relabel the symbols on  $A$  such that  $P$  can be written in the Frobenius form decomposed into primitive blocks as follows

$$P = \begin{pmatrix} \mathbf{0} & B_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & B_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & B_{h-1} \\ B_h & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{pmatrix}. \quad (11)$$

Therefore,  $P^h$  will be a primitive matrix given by

$$P^h = \begin{pmatrix} B_1.B_2.\dots.B_h & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & B_2.B_3.\dots.B_h.B_1 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & B_h.B_1.\dots.B_{h-1} \end{pmatrix} \quad (12)$$

and we have

$$\lim_{n \rightarrow \infty} \left( \frac{P^h}{\lambda_1^*} \right)^n = G^*$$

with  $G^* = \phi.\psi^t > 0$ , and  $\phi$  and  $\psi$  are the right and left Perron vectors corresponding to  $\lambda_1$  from  $P^h$ , respectively. Since,

$$p_{n,h+1} = \frac{V.P^{n,h}.W^t}{V.P^{n,h}.\mathbf{1}^t} = \frac{V.\left(\frac{P^h}{\lambda_1^*}\right)^n.W^t}{V.\left(\frac{P^h}{\lambda_1^*}\right)^n.\mathbf{1}^t}$$

and

$$p_{n,h+r+1} = \frac{V.P^{n,h+r}.W^t}{V.P^{n,h+r}.\mathbf{1}^t} = \frac{V.\left(\frac{P^h}{\lambda_1^*}\right)^n.P^r.W^t}{V.\left(\frac{P^h}{\lambda_1^*}\right)^n.P^r.\mathbf{1}^t},$$

leading to

$$\lim_{n \rightarrow \infty} p_{n,h+r+1} = \lim_{n \rightarrow \infty} \frac{V.\left(\frac{P^h}{\lambda_1^*}\right)^n.P^r.W^t}{V.\left(\frac{P^h}{\lambda_1^*}\right)^n.P^r.\mathbf{1}^t} = \frac{V.G^*.P^r.W^t}{V.G^*.P^r.\mathbf{1}^t}. \quad (13)$$

Therefore, in order to have continuity of the transtion probabilites, we need Expression (13) to be constant for all  $r = 0, \dots, h-1$ .  $\square$

In some cases, the aggregated process has finite memory, for example, it is known that if the sum of all rows of  $P$  are equal then the resulting process is Markovian (Burke and Rosenblatt, 1958). It is possible to obtain higher order Markov processes as shown by the following proposition.

**Proposition 3.2.** (a) *The process  $\mathbf{Y}$  is a Markov process with order at most  $m-1$  if  $P$  have a single non-null eigenvalue.*

(b) *The process  $\mathbf{Y}$  is a Markov process with order  $k+1$  if  $P$  have a single non-null eigenvalue and  $k$  non-null rows in its triangular form.*

*Proof.* Under the conditions of the theorem, then  $\lambda_1$  is the only non-null eigenvalue and by Theorem 1.2, p.9 (Seneta, 2006) we have

$$P^n = \lambda_1^n.\phi.\psi^t$$

for all  $n \geq m-2$ . Therefore, for any  $n \geq m-2$

$$p_{n+1} = \frac{V.P^n.W^t}{V.P^n.\mathbf{1}^t} = \frac{V.\lambda_1^n.\phi.\psi^t.W^t}{V.\lambda_1^n.\phi.\psi^t.\mathbf{1}^t} = \frac{V.G.W^t}{V.G.\mathbf{1}^t}.$$

Moreover, according to Schurs Triangularization Theorem, if all eigenvalues of  $P$  are real then  $P$  can be triangularized, that is, there exist  $Q$  and  $\Delta$ , such that

$$P = Q.\Delta.Q^t \quad \text{and} \quad P^n = Q.\Delta^n.Q^t$$

where  $\Delta$  is triangular superior and  $Q$  is orthogonal. Also,  $\Delta$  is composed by the eigenvalues in its diagonal. Suppose that  $\Delta$  has  $k$  non-null rows. Then, it is easy to see that for  $n \geq k$

$$\Delta^n = \lambda_1^{n-k}.\Delta^k.$$

Now, for  $n \geq k$

$$\begin{aligned} p_{n+1} &= \frac{V.Q.\Delta^n.Q^t.W^t}{V.Q.\Delta^n.Q^t.\mathbf{1}^t} \\ &= \frac{\lambda_1^{n-k}.V.Q.\Delta^k.Q^t.W^t}{\lambda_1^{n-k}.V.Q.\Delta^k.Q^t.\mathbf{1}^t} \\ &= \frac{V.Q.\Delta^k.Q^t.W^t}{V.Q.\Delta^k.Q^t.\mathbf{1}^t} \\ &= p_{k+1}. \end{aligned}$$

Therefore the process  $\mathbf{Y}$  is an order  $k + 1$  Markov chain.  $\square$

Notice that this Proposition does not cover the usual order 1 Markov case since  $\Delta$  cannot have zero non-null rows (the second largest eigenvalue is equal to zero, therefore at least one row of  $\Delta$  is equal to zero).

### 3.3 The reducible case

Without loss of generality consider that the transition submatriz  $P$  can be written as

$$P = \begin{pmatrix} T_{11} & T_{12} & T_{13} & \dots & T_{1c} \\ \mathbf{0} & T_{22} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & T_{33} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & T_{cc} \end{pmatrix}. \quad (14)$$

where the submatrices  $T_{22}, T_{33}, \dots, T_{cc}$  are irreducible.

The reducible case is harder to analyse in generality since the matrices  $T_{22}, T_{33}, \dots, T_{cc}$  can be periodic with different periods. However, this situation can be studied case by case using the same tools as described before.

There are some special cases, where we can get sufficient conditions for continuity.

Let  $\lambda_{1,i}$  be the Perron value of  $T_{ii}$  and define

$$\lambda_{\max} = \max_i \lambda_{1,i}. \quad (15)$$

**Definition 3.5.** *The matrices  $T_{ii}, i \in \{2, 3, \dots, c\}$  with  $\lambda_i = \lambda_{\max}$  are called Dominating Blocks.*

Let

$$\Delta = \{i \in \{2, \dots, c\} : T_{ii} \text{ is a Dominating block}\}. \quad (16)$$

**Theorem 3.2.** *If the matrix  $P$  is reducible and written as (14), then the process  $\mathbf{Y}$  is continuous if all Dominating Blocks ( $T_{ii}$  for  $i \in \Delta$ ) are*

*i either irreducible; or*

*ii are periodic with the same period  $h$  and for all  $i \in \Delta$*

$$\frac{V_i G_i (T_{ii})^r W_i^t}{V_i G_i (T_{ii})^r \mathbf{1}_i^t} = \frac{V_i G_i W_i^t}{V_i G_i \mathbf{1}_i^t}, \quad (17)$$

*for all  $r = 0, \dots, h-1$  with  $G_i = \phi_i \psi_i^t > 0$ , and  $\phi_i$  and  $\psi_i$  are the right and left Perron vectors corresponding to  $\lambda_{1,i}$  from  $T_{ii}^h$ .*

*Proof.* Analogous to Theorem 3.1. □

### 3.4 Denumerable alphabet

The theory of Perron-Frobenius can be used for countable, real, non-negative matrices. The definitions of irreducibility and period extend easily to such matrices. Moreover, we can define the Perron-value of a denumerable matrix using the following lemma.

**Lemma 3.1** (Lemma 7.1.1, Kitchens (1998)). *Let  $P$  be a countable, real, non-negative, irreducible and aperiodic matrix. For a fixed state  $i$ :*

- (i) there exists a  $k$  such that  $(P^n)_{ii} > 0$  for all  $n \geq k$ ;*
- (ii)  $(P^{n+m})_{ii} \geq (P^n)_{ii} \cdot (P^m)_{ii}$ ;*
- (iii)  $\lim_{n \rightarrow \infty} \sqrt{n} (P^n)_{ii}$  exists and equals  $\sup_n \sqrt{n} (P^n)_{ii}$ ;*
- (iv) if  $\lambda = \lim_{n \rightarrow \infty} \sqrt{n} (P^n)_{ii}$  then  $(P^n)_{ii} / \lambda^n \leq 1$  for all  $n$ .*

Then,  $\lambda$  is the *Perron value* of  $P$ . It can be infinite. For this section, we will assume it is finite. However, this is not enough to have the desired results in terms of continuity. We will need positive recurrence. This concept is well understood for stochastic matrices in terms of the time to return to the states. However, for general non-negative matrices the following definitions are necessary. Let

$$t_{ij}(0) = \delta_{ij}, \quad t_{ij}(1) = P_{ij} \quad \text{and} \quad t_{ij}(n) = (P^n)_{ij}.$$

Define the generating function as

$$T_{ij}(z) = \sum_{n=0}^{\infty} t_{ij}(n) z^n. \quad (18)$$

Now, let

$$\ell_{ij}(0) = 0, \quad \ell_{ij}(1) = P_{ij} \quad \text{and} \quad \ell_{ij}(n+1) = \sum_{r \neq i} \ell_{ir}(n) t_{rj}.$$

Observe that the radius of convergence of  $T_{ij}(z)$  is  $1/\lambda$  for all  $i, j$ .

**Definition 3.6.** *The irreducible matrix  $P$  corresponding to the generating function (18) is:*

- (i) recurrent if  $T_{ii}(1/\lambda) = \infty$ ;
- (ii) transient if  $T_{ii}(1/\lambda) < \infty$ ;
- (iii) positive recurrent if  $T_{ii}(1/\lambda) = \infty$  and  $\sum_n n \ell_{ii}(n) / \lambda^n < \infty$ ;
- (iv) null recurrent if  $T_{ii}(1/\lambda) = \infty$  and  $\sum_n n \ell_{ii}(n) / \lambda^n = \infty$ .

Notice that the definition is independent of the choice of  $i$ .

The Generalized Perron-Frobenius Theorem states the following.

**Theorem** (Theorem 7.1.3, Kitchens (1998)). *Suppose  $P$  is a countable, non-negative, irreducible, aperiodic and positive recurrent matrix. Then, there exists a Perron value  $\lambda > 0$  (assumed to be finite) such that  $\lim_{n \rightarrow \infty} P^n / \lambda^n = \phi \cdot \psi^t > 0$ , and  $\phi$  and  $\psi$  are the Perron vectors corresponding to  $\lambda$  from  $P$  and  $P^t$ , respectively.*

Therefore, we have the following result for countable Markov chains.

**Theorem 3.3.** *If the matrix  $P$  is countable, non-negative, irreducible, aperiodic and positive recurrent matrix, then the process  $\mathbf{Y}$  is continuous.*

## 4 Discussion

### 4.1 Examples

**Example 4.1. All entries positive – comparing with Harris’ rate of convergence.** Consider the Markov chain  $\mathbf{X}$  with alphabet  $A = \{1, 2, 3\}$  and transition matrix given by

$$P^{\mathbf{X}} = \begin{pmatrix} 0.10 & 0.3 & 0.60 \\ 0.20 & 0.3 & 0.50 \\ 0.05 & 0.7 & 0.25 \end{pmatrix}.$$

Notice that the process  $\mathbf{Y}$  is not Markov since  $W$  is not constant. In this case, it is easy to compute the eigenvalues and eigenvectors of  $P$  and  $P'$  corresponding to the largest eigenvalue obtaining that  $p_{\infty} = \lim_{n \rightarrow \infty} p_n = 0.132864$  and the rate of convergence

$$\beta^{\mathbf{Y}}(n) = O\left(\left(n \frac{|\lambda_2|}{\lambda_1}\right)^n\right) = O(n 0.3657281^n).$$

Notice that all entries of  $P^{\mathbf{X}}$  are positive and the continuity rate given in Harris (1955) (Eq. (6.4)) gives a bound for continuity rate as

$$\beta^{\mathbf{Y}}(n) \leq (1 - \lambda)^{n-1}$$

where

$$\lambda = \min_{i,j,k,l} \frac{p_{kj}p_{il}}{|A|^2 p_{ij}p_{kl}}.$$

In this case,  $\lambda = 0.01190476$  and

$$\beta^{\mathbf{Y}}(n) \leq (0.9880952)^{n-1}.$$

**Example 4.2.  $P$  is reducible and  $\mathbf{Y}$  is continuous.** Consider the Markov chain  $\mathbf{X}$  with alphabet  $A = \{1, 2, \dots, 7\}$  and transition matrix given by

$$P^{\mathbf{X}} = \begin{pmatrix} 0.2 & 0.3 & 0.1 & 0.2 & 0.1 & 0 & 0.1 \\ 0.2 & 0.3 & 0.2 & 0.1 & 0 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.2 & 0 & 0.3 & 0.1 & 0.2 \\ 0 & 0 & 0 & 0.4 & 0.6 & 0 & 0 \\ 0.3 & 0 & 0 & 0.2 & 0.5 & 0 & 0 \\ 0.7 & 0 & 0 & 0 & 0 & 0.3 & 0 \\ 0.6 & 0 & 0 & 0 & 0 & 0 & 0.4 \end{pmatrix}.$$

Since we have  $P$  to be reducible we have to find the dominant blocks and the maximum eigenvalue of the dominant blocks. In this case, the dominant block is

$$T_{22} = \begin{pmatrix} 0.4 & 0.6 \\ 0.2 & 0.5 \end{pmatrix}.$$

which has eigenvalues 0.8 and 0.1 and left and right normalized eigenvectors  $\psi = (0.5150787, 1.0301574)'$  and  $\phi = (0.8320503, 0.5547002)'$  respectively yielding  $p_\infty = \lim_{n \rightarrow \infty} p_n = 0.2$  and the rate of convergence is

$$\beta^{\mathbf{Y}}(n) = O\left(\left(n \frac{|\lambda_2|}{\lambda_1}\right)^n\right) = O(n(1/8)^n).$$

**Example 4.3.**  $P$  is periodic with constant sum on its columns. Consider the Markov chain  $\mathbf{X}$  with alphabet  $A = \{1, 2, \dots, m\}$  and transition matrix such that

$$P = \begin{pmatrix} \mathbf{0} & B_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & B_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & B_{h-1} \\ B_h & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{pmatrix} \quad (19)$$

with  $\sum_i P_{ij} = k$  for all  $j$ .

Notice that, in this case,  $k$  is the largest eigenvalue of  $P$  and  $\mathbf{1}$  is a left eigenvector of  $P^r$ , for all  $r$ . Therefore, Expression (13) can be written as

$$\frac{V.G^*.P^r.W^t}{V.G^*.P^r.\mathbf{1}^t} = \frac{V.\phi.\mathbf{1}.P^r.W^t}{V.\phi.\mathbf{1}.P^r.\mathbf{1}^t} = \frac{k^r V.\phi.\mathbf{1}.W^t}{k^r V.\phi.\mathbf{1}.\mathbf{1}^t} \quad (20)$$

which is constant for all  $r$ .

**Example 4.4.**  $P$  has only one non-null eigenvalue and  $Y$  is a Markov chain of finite order. Consider the Markov chain  $\mathbf{X}$  with alphabet  $A = \{1, 2, 3, 4, 5\}$  and transition matrix

$$P^X = \begin{pmatrix} 0.3 & 0.2 & 0.1 & 0.25 & 0.15 \\ 0.2 & 0.5 & 0.3 & 0.0 & 0.0 \\ 0.85 & 0.0 & 0.0 & 0.1 & 0.05 \\ 0.8 & 0.0 & 0.0 & 0.0 & 0.2 \\ 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \end{pmatrix}. \quad (21)$$

Notice that,  $\mathbf{Y}$  is not Markov of order 1 since the  $W$  is not constant. On the other hand, it is easy to see that  $P$  has only one non-null eigenvalue and its triangular form has 3 non-null rows. Therefore, by Proposition 3.2 we have an order 4 Markov chain.

In fact, it is easy to see that

$$P^k / \lambda_1^k = P^3 / \lambda_1^3$$

for all  $k \geq 3$ .

## 4.2 Relation to Gibbsianess

Fernández et al. (2011) proved that the chain  $\mathbf{Y}$  is Gibbsian (in the sense of statistical physics) if and only if we have convergence, as  $m$  and  $n$  diverge, of

$$\mathbb{P}(Y_0 = 1 | Y_{-m-1} = 1, Y_{-m}^{-1} = 0^m, Y_1^n = 0^n, Y_{n+1} = 1) =: p_{m,n}.$$

This is because Gibbsianess (in the sense of statistical physics) corresponds to the continuity of the “two-sided set of transition probabilities”, or *specification*. We do not enter further in detailed definitions, and refer the interested reader to Fernández et al. (2011), where this notion of Gibbsianess is defined.

Let  $r_{m,n}$  denote

$$\sum_{\substack{b_{-m}^{-1} \in \pi^{-1}(0_{-m}^{-1}) \\ b_1^n \in \pi^{-1}(0_1^n)}} \mathbb{P}(X_{-m}^{-1} = b_{-m}^{-1}, X_0 = 1 | X_{-m-1} = 1) \mathbb{P}(X_1^n = b_1^n, X_{n+1} = 1 | X_0 = 1)$$

and  $s_{m,n}$  denote

$$\sum_{b_{-m}^n \in \pi^{-1}(0_{-m}^n)} \mathbb{P}(X_{-m}^n = b_{-m}^n, X_{n+1} = 1 | X_{-m-1} = 1).$$

Using similar calculations as the ones performed in the proof of Proposition 3.1, it is easy to arrive at

$$p_{m,n} = \frac{r_{m,n}}{r_{m,n} + s_{m,n}}.$$

This quantity also has a simple matrix form, which is as follows,  $p_{0,0} = \frac{P_{1,1}^2}{P_{1,0}P_{0,1} + P_{1,1}^2}$ , and

$$p_{m+1,n+1} = \frac{V P^m W^t V P^n W^t}{V P^{m+n+1} \mathbf{1}^t}, \quad \text{if } m, n \geq 0.$$



The conditions for convergence, or not, of this quantity, and consequently, for Gibbsianess or non-Gibbsianess, are the same as for continuity. In particular, this means that, in the conditions we were considering in this paper, none of the image processes are at the same time non-Gibbsian and continuous, or vice-versa.

In relation to the paper of Chazottes and Ugalde (2003), it is important to notice that their notion of Gibbsianess is *not* that of statistical physics. They consider Gibbs measures in the sense of Bowen, which is slightly different. This makes complicated a direct comparison between our results and theirs.

To conclude, let us mention other recent papers with interests related to ours. As already mentioned, Verbitskiy (2011a) and Verbitskiy (2011b) prove that, if the original measure has continuous transition probabilities (that is, not necessarily Markov measure, but it could be) with summable continuity rate (that is  $\sum \beta_k < \infty$ ), then, factors of this measure still have continuous transition probabilities. Similar results were obtained concerning conservation of Bowen's Gibbsianess (Chazottes and Ugalde, 2011; Pollicott and Kempton, 2011), and of statistical physics Gibbsianess (Redig and Wang, 2010).

## References

- Burke, C. J. and Rosenblatt, M. (1958). A Markovian function of a Markov chain, *Ann. Math. Statist.* **29**: 1112–1122.
- Chazottes, J.-R. and Ugalde, E. (2003). Projection of Markov measures may be Gibbsian, *J. Statist. Phys.* **111**(5-6): 1245–1272.
- Chazottes, J.-R. and Ugalde, E. (2011). On the preservation of Gibbsianess under symbol amalgamation, *Entropy of hidden Markov processes and connections to dynamical systems*, Vol. 385 of *London Math. Soc. Lecture Note Ser.*, Cambridge Univ. Press, Cambridge, pp. 72–97.
- Darroch, J. N. and Seneta, E. (1967). On quasi-stationary distributions in absorbing continuous-time finite Markov chains, *J. Appl. Probability* **4**: 192–196.
- Fernández, R., Gallo, S. and Maillard, G. (2011). Regular  $\mathbb{G}$ -measures are not always Gibbsian, *ArXiv: 1106.4188*.
- Harris, T. E. (1955). On chains of infinite order, *Pacific J. Math.* **5**: 707–724.

- Kalikow, S. (1990). Random Markov processes and uniform martingales, *Israel J. Math.* **71**(1): 33–54.
- Kitchens, B. P. (1998). *Symbolic dynamics*, Universitext, Springer-Verlag, Berlin. One-sided, two-sided and countable state Markov shifts.
- Pollicott, M. and Kempton, T. (2011). Factors of Gibbs measures for full shifts, *Entropy of hidden Markov processes and connections to dynamical systems*, Vol. 385 of *London Math. Soc. Lecture Note Ser.*, Cambridge Univ. Press, Cambridge, pp. 246–257.
- Redig, F. and Wang, F. (2010). Transformations of one-dimensional Gibbs measures with infinite range interaction, *Markov Process. Related Fields* **16**(4): 737–752.
- Seneta, E. (2006). *Non-negative matrices and Markov chains*, Springer Series in Statistics, Springer, New York. Revised reprint of the second (1981) edition [Springer-Verlag, New York; MR0719544].
- Verbitskiy, E. (2011a). On factors of  $g$ -measures, *Indag. Math. (N.S.)* **22**(3-4): 315–329.
- Verbitskiy, E. (2011b). Thermodynamics of hidden Markov processes, *Entropy of hidden Markov processes and connections to dynamical systems*, Vol. 385 of *London Math. Soc. Lecture Note Ser.*, Cambridge Univ. Press, Cambridge, pp. 258–272.